9-1-1985

Random Cell Chi-Square Diagnostic Tests for Econometric Models: I. Introduction and Applications

Donald W.K. Andrews

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 762

RANDOM CELL CHI-SQUARE DIAGNOSTIC TESTS FOR ECONOMETRIC MODELS: I. INTRODUCTION AND APPLICATIONS

by

Donald W.K. Andrews

September 1985

ABSTRACT

This paper and its sequel, Andrews [4], extend the Pearson chi-square testing method to non-dynamic parametric econometric models, in particular, models with covariates. The present paper introduces the test and discusses a wide variety of applications. Andrews [4] establishes the asymptotic properties of the test, by extending recent probabilistic results for the weak convergence of empirical processes indexed by sets. The chi-square test that is introduced can be used to test goodness-of-fit of a parametric model, as well as to test particular aspects of the parametric model that are of interest. In the event of rejection of the null hypothesis of correct specification, the test provides information concerning the direction of departure from the null. The results allow for estimation of the parameters of the model by quite general methods. The cells used to construct the test statistic may be random and can be specified in a general form.

1. INTRODUCTION

This paper and its sequel Andrews [4], referred to as Andrews II, extend the Pearson chi-square testing method to parametric models with covariates. A chi-square test is proposed that is applicable in a wide variety of cross-sectional models, including panel data models. It can be used to test the null hypothesis that the specified parametric model is correct. This is the classical goodness-of-fit hypothesis. In addition to yielding a formal acceptance or rejection of the parametric model specification, the test provides information regarding the direction of departure from the null hypothesis, in those cases where the null hypothesis is rejected.

The chi-square test also can be used to test more specific aspects of a parametric model.

The examples we consider in this paper include testing for:

(1) overall goodness-of-fit and correct specification of various aspects of logit, probit, nested logit, and log-linear models with binomial or multinomial response variables,

(2) goodness-of-fit, heterogeneity over the space of covariates, incorrect functional form, linearity of the effect of a single covariate, omitted covariates, heteroskedasticity, exogeneity, normality of errors, and outlying errors, in normal linear regression, multivariate regression, seemingly unrelated regressions (SUR), and simultaneous equations models with or without censoring or truncation, in generalized linear models, and in normal nonlinear regression and simultaneous equations models,

(3) multivariate normality of errors in linear multivariate regression, SUR, and simultaneous equations models with or without censoring or truncation, as well as in selection and switching regression models,

(4) correct specification of the marginal distribution of unobservables, and overall goodness-of-fit in duration data models,

(5) time homogeneity, and independence of covariates and random effects in panel data models, and

(6) correct specification of the error distribution in frontier production function models.

The parametric models considered here consist of parametric families of conditional distributions of response variables given covariates. The marginal distributions of the covariates are left unrestricted--as is almost always the case in practice.

The proposed chi-square test statistic is constructed by partitioning the region in which the response variables and covariates lie into disjoint cells. Then, one calculates a quadratic form based on the difference between the observed number of outcomes in each cell, and the conditionally expected number in each cell given the observed covariates. This conditional expectation is calculated using the parametric model. If the parametric model is correct, then the observed differences are due solely to random fluctuations, and the (suitably weighted) quadratic form converges in distribution to a chi-square random variable, as the sample size increases. In contrast, if the parametric model is incorrect, then the observed differences are due to both random and systematic components in general, and the quadratic form diverges to infinity, as the sample size increases.

Three features of the chi-squared tests introduced here are noteworthy. First, the tests are applicable in models with covariates that may be discrete, continuous, or mixed. This feature differentiates the present results from the chi-square tests considered in the statistical literature. Also, it yields a broad range of applicability of the tests for econometric models.

Second, the cells may be chosen using the data itself, may have flexible shapes, and may partition the product space of the response variable and covariate spaces (as opposed to just partitioning the space of the response variables or the space of the covariates separately).

A parameter estimator is needed to decide which parametric conditional distribution to use when calculating the conditionally expected number of outcomes in each cell. The third feature of note is that this estimator can be chosen quite generally. Any regular, asymptotically normal estimator can be used. This contrasts with the Pearson chi-square test for which a rather unusual estimator, called the multinomial maximum likelihood (ML) estimator, must be used.

As will become evident below, it is the combination of the first two features above that allows one to use the proposed test statistic to test the specification of a wide range of econometric models, and to direct power against a rich array of different alternatives of interest. The flexibility afforded by these features also allows one to design the cells so that the asymptotic properties take hold with as small a sample size as possible.

The third feature above makes the chi-square test more convenient and more tractable, since one is not forced to re-estimate the model to compute some special estimator required by the test statistic. It also has some power advantages, because it allows one to choose the estimator one finds most appropriate.

The derivation of the asymptotic distribution of the chi-square test statistic is complicated by the introduction of random cells, and by the use of an estimator in the calculation of the conditional expectations. Following the lead suggested in the statistical literature (see references in Andrews II), the theory of weak convergence of empirical processes can

be exploited. We introduce a new process called the conditional empirical process, which is a stochastic process indexed by partitions. This process is shown to converge to a particular tied-down Gaussian process, dubbed a trampoline, as the sample size increases. With this result, convergence in probability of the random cells, and the continuous mapping theorem, combine to yield the desired chi-square asymptotic distribution of the test statistic.

The remainder of the paper is organized as follows: Section 2 describes and motivates the chi-square test introduced here. Section 3 discusses extensions in the statistical and econometric literature of Pearson's chi-square test, and the relationship between these extensions and the test introduced here. Section 4 considers applications of the test. Nonparametric methods of constructing cells are described. Numerous examples are presented. These examples illustrate the choice of random cells for power against particular alternatives of interest. In addition, the diagnostic information provided by the test is discussed. A brief conclusion is given in Section 5.

2. THE RANDOM CELL CHI-SQUARE TEST

This section introduces a class of random cell diagnostic tests, and motivates their consideration. Basic notation, definitions, and assumptions are presented. The specific assumptions used to establish the asymptotic chi-square null-distribution of the test statistics are left to Andrews II.

The observed sample of size n consists of the first n terms of the sequence of random vectors (Y_i, X_i), $i = 1, 2, \ldots$. Y_i denotes a vector of response variables, and X_i denotes a vector of covariates (e.g., regressors in the linear regression model).

Let $\mathbf{P}$ denote the distribution of this infinite sequence of random

vectors (rv's) under the null hypothesis. The rv's Y_i and X_i take values in $\mathcal{Y} \subset R^V$ and $\mathcal{X} \subset R^K$, respectively. We assume:

M1 $\{(Y_i, X_i) : i = 1, 2, \ldots\}$ are independent and identically distributed (iid) under $\mathbb{P}$.

For cross-sectional data the iid assumption is not as restrictive as it may appear at first. The standard method of generating observations is by random sampling. The resultant observations satisfy M1, and the conditional distribution of the response variables given the covariates may exhibit heterogeneity or heteroskedasticity. Alternatively, the observations may be generated by stratified sampling. Even in this case the iid assumption can be fulfilled, in many cases, by judicious choice of Y_i and X_i. For example, suppose there are two strata and observations are drawn from these strata in a two to one ratio. We can define the i^{th} observation (Y_i, X_i) to include three underlying observations (independently chosen), two from the first strata and one from the second. In this case, a sample of size n consists of $3n$ underlying observations. The sample is iid and the two to one ratio of the stratified sampling scheme is fulfilled.[2] Even endogenous stratified sampling, such as that considered in the negative income tax experiments by Hausman and Wise [21], can be placed in the iid framework using this method.

The null hypothesis of interest is the following:

H_0 : The conditional distribution of Y_i given X_i is in the parametric family $\{f(y|x,\theta) : \theta \in \Theta\}$, where $f(y|x,\theta)$ is a density with respect to a σ-finite measure μ, and Θ is a parameter space in R^L.

Since μ is not restricted to be Lebesgue measure, this hypothesis can

accommodate cases where the response variables are discrete, continuous, or mixed. Section 4 contains examples of all three cases. As is usual, the densities $f(y|x,\theta)$ must be reasonably smooth in θ (see Andrews II).

Let P_X denote the distribution of X_i under $\mathbb{P}$. P_X is not restricted by the null hypothesis. Thus, the covariates X_i also may be discrete, continuous, or mixed.

Let θ_0 denote the true parameter value, when the null hypothesis is true.

The general alternative hypothesis is that the conditional parametric model is incorrect, i.e., the model is misspecified. This is the standard alternative hypothesis for goodness-of-fit tests, and is usually of interest. In addition, if the alternative hypothesis is true, then one often wants not only to reject the null, but also to get some information about the direction of departure from the null of the true conditional distribution. The chi-square test statistic defined below provides such information.

Although the general alternative hypothesis is simply the negation of the null, one may be more interested in certain alternative distributions than others. This interest may arise because one has reason to believe that these distributions are especially plausible alternatives, or because of their adverse affect on the validity of other inferences that one may be interested in making. In either case, the chi-square test statistic can be designed to test specific aspects of the parametric model, and hence, to have higher power against alternatives of particular interest. This is illustrated by the examples of Section 4 below.

Having outlined the basic model and hypothesis to be tested, we now consider several items used in the definition of the chi-square test statistic. First, the test statistic introduced here relies on an estimator $\hat{\theta}$ of the unknown parameter θ_0. The estimator may be any regular, asymptotically normal estimator (see assumption E1 of Andrews II, Section 2).

Next, following the approach of Pollard [44] (who considers models without covariates), the random cells used in the construction of the test statistic are chosen from a class $\mathcal{C}$ of measurable sets in $\mathcal{X} \times \mathcal{Y}$. Not all measurable sets can be included in $\mathcal{C}$ --a restriction is necessary (see assumption RC2 of Andrews II, Section 2). Let J denote the number of cells upon which the test statistic is based. J is assumed fixed for all n. (The choice of J is discussed in Andrews II, Section 5). Let $\mathcal{G}$ be a class of partitions of $\mathcal{X} \times \mathcal{Y}$, each partition being comprised of J sets from $\mathcal{C}$. That is,

$$(2.1) \qquad \mathcal{G} = \left\{ \chi \in \mathcal{C}^J : \bigcup_{j=1}^{J} \gamma_j = \mathcal{X} \times \mathcal{Y}, \ \gamma_j \cap \gamma_k = \phi, \ \forall j \neq k \right\},$$

where γ_j and γ_k denote elements of the partition χ. For each sample size n, the J cells used to construct the test statistic are given by a random element of $\mathcal{G}$, denoted $\hat{\Gamma}$ (where $\hat{\Gamma}$ depends on n in general).

We assume that $\hat{\Gamma}$ converges in probability to some fixed partition of cells $\Gamma \in \mathcal{G}$, as the sample size increases. The precise definition of convergence in probability of random sets and partitions is given below in Andrews II, Section 2. It relies on a natural measure of closeness of sets. If the cells depend upon $\hat{\theta}$ in a continuous fashion, for example, and $\hat{\theta}$ converges in probability, then $\hat{\Gamma}$ will converge in probability to a fixed partition, as required. The above definition of the random cells, however, allows for much more general specification of the cells, than simply dependence on an estimator $\hat{\theta}$. The random cells may be defined directly as random functions of the data $\{(Y_i, X_i) : i = 1, \ldots, n\}$. Examples of numerous different methods of constructing the random cells are given in Section 4. By choosing different random cells we can construct test statistics whose power is directed against different alternatives of interest, if so desired.

We now define a new stochastic process, called the conditional empirical process, that is the basis of the chi-square test statistic. Let $P_n(\cdot)$ denote the empirical measure of the sample $\{(Y_i, X_i), i = 1, \ldots, n\}$, indexed by elements γ in $\mathcal{G}$. That is,

$$(2.2) \qquad P_n(\gamma) \equiv \frac{1}{n}\sum_{i=1}^{n} \gamma(Y_i, X_i) ,$$

where $\gamma(Y_i, X_i)$ denotes the vector of indicator functions of $(Y_i, X_i) \in \gamma_j$, for $j = 1, \ldots, J$. Let $F_n(\cdot,\theta)$ denote the conditional empirical measure constructed using the parametric conditional distribution of Y_i given X_i. That is,

$$(2.3) \qquad F_n(\gamma,\theta) \equiv \frac{1}{n}\sum_{i=1}^{n} \int_y \gamma(y, X_i) f(y|X_i, \theta) d\mu(y) \equiv \frac{1}{n}\sum_{i=1}^{n} F(\gamma, X_i, \theta) .$$

Note that the conditional empirical measure is a random function of $(X_1, \ldots, X_n)$ only.

DEFINITION: The conditional empirical process $\nu_n(\cdot,\theta)$ indexed by elements γ of $\mathcal{G}$ is defined as

$$(2.4) \qquad \nu_n(\cdot,\theta) = \sqrt{n}\Big(P_n(\cdot) - F_n(\cdot,\theta)\Big) .$$

The conditional empirical process evaluated at θ_0 and some partition γ, i.e., $\nu_n(\gamma, \theta_0)$, is a normalized measure of the difference between the observed number of observations in each cell $\gamma_1, \ldots, \gamma_J$, and the number expected in each cell according to the conditional density $f(y|x, \theta_0)$ and the observed covariates $(X_1, \ldots, X_n)$. If the null hypothesis is true, then the expected values of $P_n(\gamma)$ and $F_n(\gamma, \theta_0)$ are equal (by iterated

expectations), and so, the strong law of large numbers (SLLN) implies that $P_n(\gamma) - F_n(\gamma, \theta_0)$ converges in probability to a vector of zeroes. Further, the central limit theorem (CLT) implies that $\nu_n(\gamma, \theta_0)$ converges weakly to a multivariate normal distribution. On the other hand, if the null hypothesis is false, then $P_n(\gamma)$ and $F_n(\gamma, \theta_0)$ generally have different expectations, and so, the SLLN implies that $P_n(\gamma) - F_n(\gamma, \theta_0)$ converges in probability to a vector different from the zero vector and $\|\nu_n(\gamma, \theta_0)\|$ diverges to infinity. This difference in behavior of the conditional empirical process under the null and under the alternative is the basis for using it to construct the test statistic.

For reference, we note that the <u>standard empirical process</u> indexed by elements of $\mathcal{G}$ is defined as:

$$(2.5) \qquad \eta_n(\cdot,\theta) = \sqrt{n}\left(P_n(\cdot) - \tilde{F}(\cdot,\theta)\right),$$

where $\tilde{F}(\gamma,\theta) \equiv \int_{\mathcal{X}}\int_{\mathcal{Y}} \gamma(y,x) f(y|x,\theta)\, d\mu(y)\, dP_X(x)$. The second summand $\tilde{F}(\gamma,\theta)$ of the standard empirical process is non-stochastic, whereas the second summand of the conditional empirical process $F_n(\gamma,\theta)$ is a function of the random variables $(X_1, \ldots, X_n)$.

The Pearson chi-square statistic is based on the standard empirical process $\eta_n(\cdot,\theta)$. This test statistic <u>could</u> be used in situations where covariates are present. It generally is not used, however, for the reason stated above. Namely, it requires knowledge of the marginal distribution of the covariates (in order to calculate $\tilde{F}(\gamma,\theta)$), and such marginal distributions are very rarely specified in practice. The conditional empirical process $\nu_n(\cdot,\theta)$ does not depend on the unspecified marginal distribution of the covariates. Hence, it is the appropriate basis for a chi-square test statistic in models with covariates.[3]

Since θ_0 and Γ are unknown, $\nu_n(\Gamma, \theta_0)$ cannot be used to form a test. Instead, the estimator $\hat{\theta}$ and the random cells $\hat{\Gamma}$ are employed, and we consider a test based on the statistic $\nu_n(\hat{\Gamma},\hat{\theta})$.

If $\nu_n(\hat{\Gamma},\hat{\theta})$ is sufficiently different from the zero vector, then the test should reject the null hypothesis. How does one measure difference from the zero vector? A natural method is to choose the critical region such that, asymptotically, its boundary is a constant probability density contour. This method has the advantage that a local perturbation of the boundary of the critical region cannot reduce the volume of the critical region while maintaining the same significance level. If $\nu_n(\hat{\Gamma},\hat{\theta})$ has asymptotic normal distribution, then this criterion leads to defining the test statistic as a quadratic form in $\nu_n(\hat{\Gamma},\hat{\theta})$ with weighting matrix given by a consistent estimator of a g-inverse of its asymptotic covariance matrix.

The limit distribution of $\nu_n(\hat{\Gamma},\hat{\theta})$ does not follow directly from the CLT, because the cells $\hat{\Gamma}$ and the estimator $\hat{\theta}$ are random. Nevertheless, using the weak convergence of $\nu_n(\cdot, \theta_0)$ as a process indexed by partitions $\gamma \in \mathcal{G}$, $\nu_n(\hat{\Gamma},\hat{\theta})$ can be shown to have asymptotic normal distribution under the null, with mean vector zero and some covariance matrix Σ_0 (see Andrews II, Section 3). Furthermore, a consistent estimator $\hat{W}$ of some generalized inverse of Σ_0 can be constructed. Hence, we define the test statistic as follows:

DEFINITION: The random cell chi-square test statistic is given by

$$(2.6) \qquad X_n^2(\hat{\Gamma},\hat{\theta}) = \nu_n(\hat{\Gamma},\hat{\theta})'\hat{W}\nu_n(\hat{\Gamma},\hat{\theta}) \ .$$

The chi-square test rejects the null hypothesis if $X_n^2(\hat{\Gamma},\hat{\theta})$ is sufficiently large. Given the results stated above, the test statistic $X_n^2(\hat{\Gamma},\hat{\theta})$

has limiting chi-square distribution under the null with degrees of freedom given by the rank of Σ_0 (denoted $rk[\Sigma_0]$). Generally, the rank of Σ_0 is J-1. If a special estimator called a minimum chi-square estimator (defined in Andrews II, Section 3) is employed, however, then the rank of Σ_0 is J-1-L (where L is the dimension of θ_0). This estimator is unnatural and unlikely to be used, except in the case where the model has response variables and covariates that take on at most a finite number of different values, and the cells completely cross-classify the response variables and the covariates. In this case, ML estimation and minimum chi-square estimation are asymptotically equivalent. This is the situation considered by McFadden [31] in the particular case where the conditional parametric model is multinomial logit.

A second special case where $rk[\Sigma_0] \neq J-1$, arises when no covariates are present and the parameter θ is estimated by the multinomial ML estimator. This estimator maximizes the likelihood of the data after grouping the data into J cells. In this case, the chi-square statistic of (2.6) reduces to the classical Pearson chi-square statistic (for appropriate choice of $\hat{W}$), and the rank of Σ_0 is given by the familiar expression J-1-L.

If the parameter θ is not estimated by the multinomial ML estimator (or some asymptotic equivalent)[4] in this case, then the Pearson chi-square statistic has a limit distribution under the null that is an intractable mixture of chi-square distributions, see Chernoff and Lehmann [8]. In contrast, with the weighting matrix considered here, the test statistic of (2.6) adjusts appropriately with the form of the estimation procedure, and has chi-square limit distribution under the null, with degrees of freedom given by $rk[\Sigma_0]$. In this case, $rk[\Sigma_0]$ equals J-1 in general, rather than J-1-L.

As mentioned above, the random cell chi-square test provides information about the direction of departure from the null hypothesis, in the event that the null hypothesis is rejected. This information is garnered by considering the pattern of deviations of the observed number, from the expected number, of outcomes in each cell. That is, one compares $P_{nj}(\hat{\Gamma})$ with $F_{nj}(\hat{\Gamma}, \hat{\theta}_n)$, for $j = 1, \ldots, J$, where $P_{nj}(\cdot)$ and $F_{nj}(\cdot,\cdot)$ denote the j^{th} elements of $P_n(\cdot)$ and $F_n(\cdot,\cdot)$, respectively.

Since different cells may have different expectations, it is appropriate to re-scale these numbers initially. This rescaling can be carried out using an estimator $\hat{\Sigma}$ of the asymptotic covariance matrix Σ_0 of $\nu_n(\hat{\Gamma},\hat{\theta})$. In particular, consider the J deviations $\sqrt{n}(P_{nj}(\hat{\Gamma}) - F_{nj}(\hat{\Gamma},\hat{\theta}))/\hat{\sigma}_{jj}$ $(\equiv \nu_{nj}(\hat{\Gamma},\hat{\theta})/\hat{\sigma}_{jj})$, for $j = 1, \ldots, J$, where $\hat{\sigma}^2_{jj}$ denotes the $(j,j)^{th}$ element of $\hat{\Sigma}$. These deviations have standard normal asymptotic distributions. Hence, one can calculate easily which deviations are individually statistically significantly different from zero, and one can compare in a straightforward fashion the patterns in the signs and magnitudes of the deviations from one cell to another.

The interpretation of these patterns in terms of the direction of departure from the null hypothesis is straightforward in some cases, but less so in others. The examples of Section 4 below illustrate how these interpretations can be made.

3. EXTENSIONS OF PEARSON'S CHI-SQUARE TEST

This section describes the developments and extensions of Pearson's chi-square test that have been made in the statistical and econometric literature. It also discusses the relationship of these extensions with the test introduced in Section 2 above.

Pearson's [42] chi-square test was the first method devised for testing the appropriateness of a specified statistical model. Its introduction was an important step in the development of statistical methodology, and in some sense, ushered in the age of statistical inference. In fact, the historical significance of the Pearson chi-square test was recognized recently when Science magazine included it in a list of the "20 discoveries in science, technology, and medicine that we believe rank among this century's most significant historical developments in any field."[5]

Pearson's test applies when one has a sample of n iid random variables or vectors and a parametric family of distributions that one believes contains the true distribution of the rv's. Let $\{Y_i : i = 1, \ldots, n\}$ denote the sample of rv's and $\{F_\theta : \theta \in \Theta\}$ denote the parametric family of distributions. Pearson considered the goodness-of-fit problem of testing the composite null hypothesis $H_0 : F = F_\theta$, for some $\theta \in \Theta$, against the alternative hypothesis $H_1 : F \neq F_\theta$, for any $\theta \in \Theta$, where F denotes the distribution of Y_i. He suggested partitioning the range space of Y_i into J cells, and forming a test based on the difference between the number of observations in each cell, and the expected number in each cell when the null hypothesis is true. Since the expected number in each cell depends on the value of θ, he used an estimator $\hat{\theta}$ of θ when calculating his test statistic. (He was not specific regarding the choice of estimator $\hat{\theta}$.) In particular, Pearson proposed the following statistic:

$$(3.1)\quad X_n^2(\hat{\theta}) \equiv \sum_{j=1}^{J} \frac{[\text{observed}(\gamma_j) - \hat{\text{expected}}(\gamma_j)]^2}{\hat{\text{expected}}(\gamma_j)} = \sum_{j=1}^{J} \frac{\left[\sum_{i=1}^{n} 1(Y_i \in \gamma_j) - n \cdot F_{\hat{\theta}}(\gamma_j)\right]^2}{n \cdot F_{\hat{\theta}}(\gamma_j)}$$

where $\{\gamma_j : j = 1, \ldots, J\}$ are the J cells that partition the range space of Y_i, $1(Y_i \in \gamma_j)$ is the indicator function that equals one when $Y_i \in \gamma_j$ and zero otherwise, and $F_{\hat{\theta}}(\gamma_j)$ is the probability under $F_{\hat{\theta}}$ that $Y_i \in \gamma_j$.

Pearson claimed that the distribution of $X_n^2(\hat{\theta})$ in large samples is approximately chi-squared with $J-1$ degrees of freedom--just as it would be if θ_0 was known, rather than estimated. His claim turned out to be incorrect. Fisher [13] proved that if the multinomial maximum likelihood estimator is used to estimate θ, then the actual asymptotic distribution of $X_n^2(\hat{\theta})$ is a chi-square distribution with $J-1-L$ degrees of freedom, where L is the dimension of θ_0.[6] (The multinomial ML estimator is the estimator that maximizes $\prod_{i=1}^{n} \prod_{j=1}^{J} F_\theta(\gamma_j)^{1(Y_j \in \gamma_j)}$ over $\theta \in \Theta$.)

Fisher's proof was improved by Cramer [10], Birch [5], and Rao [48] in terms of rigor and breadth of applicability. Neyman [39] showed that the Pearson chi-square test (with the multinomial ML estimator) is consistent against all alternatives. In one of the first uses of local power, Eisenhart [12] found its asymptotic distribution under local alternatives to be non-central chi-square.

Unless the original model is multinomial, the use of the multinomial ML estimator is unnatural and inconvenient. It is asymptotically inefficient and requires the evaluation of J integrals in each iteration of the solution of the maximization problem. In consequence, Chernoff and Lehmann [8] investigated the asymptotic distribution of Pearson's chi-square statistic when the standard ML estimator is employed. They found the limit distribution to be a mixture of independent chi-square rv's under the null, with

the mixture depending on the unknown parameter θ_0 . Hence, the use of the standard ML estimator is problematic.

Roy [51] and Watson [54, 55, 56] suggested a clever method of circumventing the Chernoff-Lehmann problem by making use of data-dependent random cells. Their method only works, however, when θ consists of location-scale parameters.

A second method of circumventing the Chernoff-Lehmann problem was proposed by Nikulin [40] and Rao and Robson [49]. This method is not restricted to location-scale problems, and hence, is of interest for the general problems considered here.

The Pearson chi-square statistic is a quadratic form in the J-vector $v_n(\hat{\theta})$ with elements $v_{nj}(\hat{\theta}) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [1(Y_i \in \gamma_j) - F_{\hat{\theta}}(\gamma_j)]$, $j = 1, \ldots, J$, and with $J \times J$ weighting matrix $W(\hat{\theta}) \equiv \text{diag}[1/F_{\hat{\theta}}(\gamma_1), \ldots, 1/F_{\hat{\theta}}(\gamma_J)]$. For most estimators $\hat{\theta}$, $v_n(\hat{\theta})$ has an asymptotic normal distribution under the null hypothesis. In the special case where $\hat{\theta}$ is the multinomial ML estimator, the weight matrix evaluated at the true θ_0 , i.e., $W(\theta_0)$, is a generalized inverse of the asymptotic covariance matrix of $v_n(\hat{\theta})$. In consequence, the quadratic form given by the Pearson chi-square statistic has asymptotic chi-square distribution.

For other estimators, including the standard ML estimator, $v_n(\hat{\theta})$ generally has a different asymptotic covariance matrix, for which the Pearson weighting matrix $W(\theta_0)$ is not a generalized inverse. In consequence, the Pearson chi-square test is not asymptotically chi-squared for arbitrary estimation methods.

Nikulin [40] and Rao and Robson [49] suggested the natural solution to the problem. Instead of considering the quadratic form used by Pearson, they proposed a quadratic form whose weighting matrix is a consistent estimator

of a generalized inverse of the asymptotic covariance matrix of $v_n(\hat{\theta})$. The resulting quadratic form then has a chi-square asymptotic distribution with degrees of freedom given by the rank of the asymptotic covariance matrix of $v_n(\hat{\theta})$. This method is adopted in the present paper.

The introduction of random cells in the theoretical work of Roy and Watson was instigated not only by the desire to circumvent the Chernoff-Lehmann problem, but also to reflect the procedures actually used in practice. It is fairly common in practice to use the data, to a greater or lesser extent, when choosing the cells to be used in chi-square tests (e.g., see Kendal and Stuart [27, p. 448]). The potential advantages of doing so include: improvement of power properties, more rapid convergence to the limiting chi-square distribution under the null, and the ability to construct tests that are invariant under various transformations of the data. Furthermore, as the examples of Section 4 illustrate, the use of random cells allows one to construct chi-square tests with high power directed toward particular alternatives of interest. This feature of random cell chi-square tests is exploited greatly in the present paper and Andrews II.

Watson presented only a sketch of the proof of his results for tests with random cells. In the early seventies, however, Chibisov [9] and Moore [33] independently established results (including Watson's) for Pearson chi-square tests with data-dependent random cells, using the theory of weak convergence. Moore and Spruill [35] extended these results, but still required quite restrictive assumptions on (i) the shape of the cells, viz., rectangular with edges parallel to the coordinate axis, and (ii) the way in which the data could determine the cells, viz., through a finite dimensional parameter estimator. These restrictions were relaxed greatly in a very elegant paper by Pollard [44], whose approach is followed here. Pollard used the weak

convergence results of Dudley [11] to obtain his generalizations. Tauchen [53] also has relaxed Moore and Spruill's assumptions on cell shapes. Tauchen's results, however, require that the cells depend on the data only through the maximum likelihood estimator.

All of the results discussed above (in this section) assume that a parametric model is given that completely specifies the distribution of the data up to an unknown parameter. Few econometric models fit into this framework. For models with covariates, one usually specifies a conditional parametric model for the response variable(s), given the value of the covariates. With this approach, one avoids placing restrictive and unnecessary assumptions on the marginal distribution of the covariates.

To apply any of the above chi-square tests to a conditional parametric model, one would have to specify a marginal distribution for the covariates. Rejection of the null hypothesis, in this case, may reflect incorrect choice of the marginal distribution of the covariates, rather than misspecification of the conditional parametric family, which is of primary interest. Hence, this solution is not satisfactory.

Two chi-square tests suggested in the econometrics literature avoid the specification of the marginal distribution of the covariates. The first is by McFadden [31], and is designed for the multinomial logit model when the covariates take on only a finite number of different values. As McFadden recognized, this is a very special case, with only a limited number of applications in econometrics.

A second test has been suggested by Heckman [22]. Heckman considers a parametric model for the conditional distribution of a univariate, absolutely continuous response variable, given a vector of covariates.[7] He partitions the range space of the response variable, and considers the

deviations of the observed number of outcomes in each cell from the conditionally expected number, given the observed sequence of covariates. This conditional expectation is calculated using the postulated conditional parametric model evaluated at the ML estimator. Heckman then constructs a quadratic form in these deviations, with weighting matrix that adjusts for the estimation of θ (as in Nikulin [40] and Rao and Robson [49]). This quadratic form is his test statistic, and it has chi-square asymptotic distribution under the null.

Since Heckman's test statistic does not require the marginal distribution of the covariates to be specified, it has fairly wide applicability. It is limited, however, by the following factors: It requires fixed cells, rather than data dependent random cells; the cells are allowed to partition only the range space of the response variable, rather than the product space of the response variable and covariates; the cells must be of a specific shape, viz., intervals; the response variable is restricted to being univariate and absolutely continuous; and the parameter vector θ must be estimated by maximum likelihood.[8]

The test statistic proposed in Section 2 above is not subject to these restrictions. It utilizes the random cell approach of Chibisov [9], Moore [33], Moore and Spruill [35], and Pollard [44], but allows for covariates, as in McFadden [31] and Heckman [22]. The random cells may be of general shape and may partition the product space of the response variables and covariates. The response variables and covariates may have continuous, discrete, or mixed distributions. The test statistic uses the adjusted weighting matrix approach introduced by Nikulin [40] and Rao and Robson [49]. Hence, arbitrary asymptotically normal estimators can be employed. As the examples of Section 4 illustrate, the above conditions give the test considerable

flexibility and wide applicability.

The results given here and in Andrews II extend those of Moore and Spruill [35], Tauchen [53], and Pollard [44] by allowing covariates. They also extend [35] and [53] by allowing more general cell shapes, and [53] and [44] by allowing more general estimation methods. Finally, they extend Heckman's [22] results by eliminating the above-mentioned restrictions on his test.

In special cases, the test statistic proposed here reduces to (i) Pearson's chi-square statistic, (ii) Moore and Spruill's chi-square statistic, (iii) McFadden's chi-square statistic, and (iv) Heckman's chi-square statistic.

4. APPLICATIONS

In this section, we discuss applications of the random cell chi-square test. We begin by considering "nonparametric" methods of constructing random cells. These methods are particularly appropriate for testing overall goodness-of-fit. Then, we discuss methods for testing a wide variety of different aspects of model specification, using an array of different examples.

4.1. Nonparametric Partitioning

Nonparametric partitioning methods are ones that do not rely on the specified conditional parametric model to form cells. These methods can be used when constructing general goodness-of-fit tests, and can be used, in part, when constructing tests of various specific aspects of a model.

For goodness-of-fit tests, four basic nonparametric partitioning methods are possible: (1) group the response variables and covariates together, and nonparametrically partition $\mathcal{Y} \times \mathcal{X}$, (2) nonparametrically partition $\mathcal{Y}$

and $\mathcal{X}$ separately, and form cross-product cells in $\mathcal{Y} \times \mathcal{X}$, (3) first partition $\mathcal{X}$, and then separately partition $\mathcal{Y}$ for each $\mathcal{X}$ cell, and (4) first partition $\mathcal{Y}$, and then separately partition $\mathcal{X}$ for each $\mathcal{Y}$ cell.

Method (1) is the least structured approach. It disregards the investigator's distinction between response variables and covariates. Method (2) has the distinct advantage that one can use the normalized cell deviations $\nu_{nj}(\hat{\Gamma},\hat{\theta})/\hat{\sigma}_{jj}$, $j = 1, \ldots, J$, to see which regions in $\mathcal{X}$, or in $\mathcal{Y}$, are inadequately modelled by the conditional parametric distributions. This method has the disadvantage, however, that it may create numerous low probability cells with few observations. Methods (3) and (4) avoid this problem. Method (3) allows one to see which regions in $\mathcal{X}$ are inadequately modelled by the conditional parametric distributions, using the normalized cell deviations. Method (4) does the same for regions in $\mathcal{Y}$. These two methods are probably the most useful, in general, for testing goodness-of-fit. The choice between the two depends on which set of direction of departure diagnostics one finds of greater interest.

When testing particular aspects of a model, nonparametric partitioning methods often are useful at one stage in the formation of cells. For example, one might nonparametrically partition $\mathcal{X}$, and then partition $\mathcal{Y}$ using a parametric method that directs power against specific alternatives of interest. The reverse procedure, of course, also arises. Alternatively, one might nonparametrically partition $\mathcal{X}$ based on a single, or small number of, covariates, and then partition $\mathcal{Y}$ in some fashion.

We now describe a number of nonparametric partitioning methods. First, consider partitioning the space of a single, real-valued variable, say w, based on n observations of the variable, into a predetermined number of cells. A simple method, that has been used for some time (see [54, 55, 56]),

is to calculate the sample mean $\overline{W}$ and standard deviation s_W of the variable, and form cells centered at the mean, with width proportional to the standard deviation. That is, take the cells as follows:

(4.1) $[\overline{W} + b \cdot c \cdot s_W, \overline{W} + (b+1) \cdot c \cdot s_W)$, $(\overline{W} - (b+1) \cdot c \cdot s_W, \overline{W} - b \cdot c \cdot s_W]$, $[\overline{W} + B \cdot c \cdot s_W, \infty)$, and $(-\infty, \overline{W} - B \cdot c \cdot s_W]$, for $b = 1, \ldots, B$, where c is a constant.

A second method is to use the k-means clustering procedure, see [17, 45]. This procedure forms k cells that minimize the within cell sum of squares for the n observations. Equivalently, it finds k optimal centers that minimize the sum of squared deviations of the observations from the nearest centers. To each center there corresponds a cell that includes all points that are closer to it than to any other center. Algorithms are available for carrying out the k-means clustering procedure, see [17, 50, 52].

A third method of partitioning a real-valued variable is to form cells to have an equal number of observations per cell, with the cell boundaries given by the mid-point between observations that fall into two different groups. Additional partitioning methods are given in the literature on cluster analysis (e.g., see [17, 50, 52]).

Next, consider the general problem of partitioning a vector-valued variable, say $\underset{\sim}{W}$, based on n observations of the variable, into a predetermined number of cells. The first method discussed above can be generalized to vector-valued variables by taking the cell boundaries to be concentric ellipses centered at the sample mean $\underset{\sim}{\overline{W}}$ and with shape determined by the sample covariance matrix S_W. In particular, for constants $a_1, a_2, \ldots, a_B$ with $a_1 = 0$ and $a_B = \infty$, form cells given by

(4.2) $$\{\underset{\sim}{W} : a_b \leq (\underset{\sim}{W} - \underset{\sim}{\overline{W}})' S_W^{-1} (\underset{\sim}{W} - \underset{\sim}{\overline{W}}) < a_{b+1}\}, \quad \text{for } b = 1, \ldots, B.$$

If desired, these cells can be split along the axes of the ellipses. These cells are particularly easy to construct.

The k-means partitioning method generalizes to vector-valued variables without change. This method forms cells by grouping points that are close to each other in Euclidean distance. In contrast, the k-means method groups points that are approximately equidistant from the center of the observations (using a metric based on the sample covariance matrix). It should be pointed out that the k-means procedure is not invariant to scale changes in the variables. Hence, appropriate scaling must be carried out before applying a k-means algorithm, see [17].

The third method discussed above, for partitioning a real-valued variable, does not generalize in a straightforward fashion to vector-valued variables. The reason is that there is no natural ordering in multi-dimensional spaces, and so, there are many different ways of forming cells such that each cell contains the same number of observations. On the other hand, numerous additional clustering procedures apply for vector-valued variables, see [17, 50, 52].

The final nonparametric partitioning method we consider is one in which a vector-valued variable, say of dimension D, is reduced to a real-valued variable, and then one of the partitioning methods discussed above for real-valued variables is applied. A reduction to a real-valued variable can be carried out by considering the first principal component of the D n-vectors of observations. Other methods of reduction to a real-valued variable also can be used.

4.2. *Examples*

We now consider a number of different econometric models, and illustrate how random cell chi-square tests can be applied in these models for various purposes. The examples illustrate the flexibility and general applicability of the tests.

To avoid repetition, we note here that in any of the models considered below, the parameters may be subject to restrictions of any sort, linear or nonlinear. For econometric applications, this greatly increases the range of applicability of the tests.

We also mention here that alternative tests exist for some, but certainly not all, of the testing situations considered below. The chi-square tests discussed here are of a different nature than most existing tests in the literature, and hence, normally have maximum power against different alternative distributions, than existing tests. Thus, we do not advocate supplanting existing tests by random cell chi-square tests. Rather, we argue that they may be useful, even when alternative tests exist, depending upon the alternative distributions of most interest. Further, in a variety of circumstances, random cell chi-square tests provide tests where no other tests are available. They also provide valuable direction of departure diagnostics.

1. Categorical Response Models--references: Amemiya [1], McFadden [32], Maddala [28, Chs. 3, 5], Nerlove and Press [38], Cavanaugh [6].

The models considered here include binomial and multinomial logit, probit, nested logit, modified logit, generalized extreme value (GEV), and log-linear models. The space of the response variable $\mathcal{Y}$ is partitioned into a finite number of classes, by the nature of the model itself. Two basic methods arise for partitioning $\mathcal{Y} \times \mathcal{X}$: (1) Partition $\mathcal{X}$, and then

form cells in $\mathcal{Y} \times \mathcal{X}$ by taking the product of cells in $\mathcal{X}$ with those in $\mathcal{Y}$, see Figure 1. (2) For each cell in $\mathcal{Y}$ (i.e., each value of Y), partition $\mathcal{X}$ by some procedure, see Figure 2. This method has the advantage that it can be designed to avoid cells with very low probability.

The best procedure for partitioning $\mathcal{X}$, for either of the two methods above, depends on which alternatives are of greatest interest. The possibilities include the following: (i) Nonparametric. Any of the nonparametric methods discussed above can be utilized. This approach is appropriate for constructing general goodness-of-fit tests. Such tests measure the goodness-of fit of the conditional parametric model for each cell of covariates, and so, also indicate whether there exists heterogeneity over the space of covariates. (ii) Value of $X_i'\hat{\beta}$ (where the conditional distributional of Y_i given X_i depends on X_i only through $X_i'\beta_0$, and $\hat{\beta}$ is an estimator of β_0). For example, the cells can be formed using the sample mean of $X_i'\hat{\beta}$ plus or minus several multiplicative factors of its standard deviation, as in (4.1). In the case of method (1) above, the sample mean and standard deviation are calculated using the whole sample. With method (2), they are calculated separately for the observations in each $\mathcal{Y}$ cell. Tests based on the value of $X_i'\hat{\beta}$ partition $\mathcal{X}$ such that the variation of within cell probabilities is small. With this method, one can determine if high or low probability outcomes are systematically over- or under-estimated. (iii) Value of Some Power(s) of $X_i'\hat{\beta}$. By analogy with the RESET test for linear regression, one can test the linear functional form $X_i'\beta_0$ by partitioning based on $(X_i'\hat{\beta})^2$, $(X_i'\hat{\beta})^3$, ... (iv) Value of a Single Covariate. By partitioning on a single covariate, one can test whether linearity in this variable adequately captures its effect. (v) Omitted variable. A variable addition test can be carried out by appending a variable Z, which does not appear

in the conditional density $f(y|x,\theta)$, to the vector of covariates. One dimension of $\mathcal{X}$ corresponds to Z, then, and one can partition $\mathcal{X}$ based on values of Z, see Figure 3. The resultant test indicates whether the omitted variable Z actually belongs in the parametric model.

The statistics $v_{nj}(\hat{\Gamma},\hat{\theta})/\hat{\sigma}_{jj}$, for $j = 1, \ldots, J$ are the normalized deviations of observed from expected cell counts (see Section 2). By inspecting the pattern of these statistics one can see which cells are over- or under-estimated. For example, with test (ii), one can see if high or low probability cells are over- or under-estimated; with test (iv), one can see if large or small values of the covariate of interest are over- or under-estimated; and with test (v), one can see how the omitted variable is associated with cell deviations. In the case of nonparametric partitioning (i), the cell deviations can be used to get an idea of the regions in $\mathcal{X}$ where the model is adequate, and the regions where it is less adequate or inadequate.

2. Structured Categorized Response Models

(a) Count Models (Poisson Regression)--references: Holland [25], Hausman, Hall, and Griliches [19], Hausman, Ostro, and Wise [20].

The most widely used count data model is the Poisson regression model. In this model, Y_i takes values $0, 1, 2, \ldots$. Some grouping of the response variable is necessary to restrict consideration to a finite number of cells. As with all models considered here, this grouping can be done using the data. Thus, we can form groups in such a manner that no group is too large or too small in terms of its probability of occurrence.

Three basic methods can be used to partition $\mathcal{Y} \times \mathcal{X}$: (1) Partition $\mathcal{Y}$ and $\mathcal{X}$ separately, and form cross-product cells in $\mathcal{Y} \times \mathcal{X}$. (2) Partition $\mathcal{Y}$, then for each $\mathcal{Y}$ cell partition $\mathcal{X}$. This method is appropriate

if one is particularly interested in certain outcomes of the response variable, perhaps low or high count outcomes. (3) Partition $\mathcal{X}$, then for each $\mathcal{X}$ cell partition $\mathcal{Y}$. This method is appropriate if one is interested in assessing the relative adequacy of the parametric model over different regions in $\mathcal{X}$. The partitioning of $\mathcal{Y}$ for a given cell in $\mathcal{X}$ may make use of an estimated value of θ. The power properties of the test may be enhanced by using an estimator of θ based on all the data except those falling into the $\mathcal{X}$ cell in question. The reason is that there is no fitting of the estimated value of θ to the observations under consideration. On the other hand, this procedure is more burdensome computationally, than just using the estimator of θ based on all of the data.

When partitioning $\mathcal{Y}$, two criteria arise. First, one may want to create a finer partition for counts of greater interest. Second, one usually wants to avoid creating cells with too low probabilities. When partitioning $\mathcal{X}$, the procedures (i)-(v) of Example 1 can be used to test the same items of interest as discussed there. The cell deviations can be used analogously to obtain information regarding direction of departure from the null hypothesis.

(b) Ordered-Response and Sequential-Response Models--reference: Maddala [28, Ch. 2].

These models have a finite number of outcomes for the response variable, and hence, $\mathcal{Y}$ does not need to be partitioned. $\mathcal{Y} \times \mathcal{X}$ can be partitioned in the same manner as in Example 1 for unstructured categorical response models. Alternatively, the choice of cells may reflect the structure on the categories. For example, in a sequential response model one might test for adequacy of the model in describing some particular stage of the sequential process in question.

The natural ordering or sequencing of the outcomes of the response variable can be used to help interpret the pattern of cell deviations, and hence, to help identify the direction of departure from the null hypothesis, when the test rejects.

3. Normal Linear Regression Models

The model considered here is the linear regression model with normally distributed errors. As mentioned above, the parameters can have restrictions of any sort, linear or nonlinear, so many models usually considered to be nonlinear regression models actually fall into this category.

We consider several different aspects of the model that one may be interested in testing:

(i) Overall goodness-of-fit. Partition $\mathcal{Y} \times \mathcal{X}$ non-parametrically by any of the methods discussed in Section 4.1 above. The resultant test statistic may be a more interesting goodness-of-fit statistic than the conventional R^2 statistic, because the latter simply measures the degree of association between the covariates and the response variable, as opposed to measuring the fit of the parametric model. In many cases in econometrics, a high degree of association is known to exist a priori, and the R^2 measure is not particularly relevant.

(ii) Heterogeneity over $\mathcal{X}$. Of interest here is the question of whether the regression function is different in different regions of $\mathcal{X}$. A test can be constructed by partitioning $\mathcal{X}$ nonparametrically, or based on values of $X_i'\hat{\beta}$ (where $X_i'\beta_0$ is the regression function), and then partitioning $\mathcal{Y}$ for each $\mathcal{X}$ cell according to the values of the residuals, see Figure 4. As in Example 2, if the estimator of θ that is used in partitioning $\mathcal{Y}$ for a given $\mathcal{X}$ cell is calculated without using the observations in the $\mathcal{X}$ cell, the power of the test may be enhanced.

(iii) Incorrect Functional Form. An analogue of the RESET test can be constructed by partitioning $\mathcal{X}$ based on powers of $X_1'\hat{\beta}$, and then partitioning $\mathcal{Y}$ for each $\mathcal{X}$ cell based on values of the residuals, perhaps into equiprobable cells. To minimize power against incorrect specification of the error distribution as normal, $\mathcal{Y}$ can be partitioned into just two cells, one above the estimated median regression line and one below the estimated median regression line. For this choice of cells, the normality assumption only is used in estimating one component of the weighting matrix of the chi-square statistic. Future research will consider non-parametric estimation of this component, so that one can test for incorrect functional form, and various other aspects considered here (such as (ii), (iv), (v), (vi), and (vii)), without assuming normality (or some other distribution for the errors).

(iv) Linearity of the Effect of a Single Regressor. In this case we are interested in the correctness of the functional form with respect to a single regressor (covariate). As in Example 1(iv), partition $\mathcal{X}$ based on the covariate in question, and then partition $\mathcal{Y}$ based on values of the residuals. $\mathcal{X}$ can be partitioned using its sample mean plus or minus several multiplicative factors of its standard deviation (see (4.1) above), or by some other nonparametric method.

(v) Omitted Regressor Variable. To test whether a variable Z has been erroneously omitted from the regression function, add the variable to the vector of covariates and partition $\mathcal{X}$ based on its values, see Figure 3, and then partition $\mathcal{Y}$ based on the values of the residuals.

(vi) Heteroskedasticity. Depending upon the information available regarding the aspect of the covariates that is related to the variance of the errors, one can partition $\mathcal{X}$ as in (ii), (iii), (iv), or (v) above, and then

partition $\mathcal{Y}$ for each $\mathcal{X}$ cell based on the absolute values of the residuals. By partitioning based on absolute values of the residuals, power is directed towards heteroskedasticity, rather than incorrect functional form, omitted variables, etc. If little or no information is available to guide the partitioning of $\mathcal{X}$, then a nonparametric partitioning can be used.

(vii) Exogeneity. Suppose one wishes to test whether a regressor X_{1i} is independent of the errors. A test can be constructed by estimating θ using an instrumental variables (IV) estimator, partitioning $\mathcal{X}$ based on the values of X_{1i}, and then partitioning $\mathcal{Y}$ for each $\mathcal{X}$ cell according to the values of the IV residuals.

(viii) Normality of Errors. To test for normality of the error distribution, partition $\mathcal{Y} \times \mathcal{X}$ according to the values of the residuals alone, see Figure 5. Do not partition on $\mathcal{X}$ first. The same method can be used to test for any specified distribution of errors, the only difference arises in the calculation of the expected number of observations in each cell.

(ix) Outlying Errors. To test for outliers in the errors, estimate the parameter vector robustly, and partition $\mathcal{Y} \times \mathcal{X}$ according to the values of the residuals alone. The partitioning should be crude near the regression line, and more fine in the tail areas. Robust estimation of the parameters is necessary, because least squares estimation may hide outliers in the process of minimizing the sum of squared residuals.

For the tests (i)-(ix) above, one may design the cells such that the test statistic is invariant under various transformations of the data and the parameters, such as changes in their units. Only with data-dependent random cells can such invariance properties be achieved.

By looking at the normalized cell deviations $\nu_{nj}(\hat{\Gamma},\hat{\theta})/\hat{\sigma}_{jj}$, $j = 1, \ldots, J$, one can detect the direction of departure from the null

hypothesis, in much the same fashion as in Example 1. For example, with test (ii), one can see for which regions of $\mathcal{X}$ the model is inadequate. With test (v), one can detect directions of curvature of the regression function in different regions of $\mathcal{X}$. With test (vi), one can find regressors that are related to the error variances. With test (viii), the cell deviations show if the normal distribution has tails that are too thin or too fat, as well as whether the true error distribution is skewed or not. Finally, with test (ix), cell deviations of cells in the tails exhibit the character of outliers that may be present.

Numerous tests already exist in the literature for testing aspects (i)-(ix) of the linear regression model. Many of the existing tests are of a single basic type, viz., variable addition, where the added variable often is continuous in nature, see Pagan [41]. These tests are of a different nature than the random cell chi-square tests considered here. In consequence, their power properties usually can be expected to be superior for some alternatives, but inferior for others.

4. Censored and Truncated Normal Linear Regression Models--references: Amemiya [3], Maddala [28, Ch. 6]

The models considered here are censored and truncated normal linear regression models with single or double censoring/truncation points. The same items of interest arise here as in the standard regression model, viz., tests (i)-(ix) of Example 3. The satisfaction of certain assumptions is more crucial in the present context, however, than in the standard linear regression model. For example, the quasi-maximum likelihood estimator (based on the assumption of iid normal errors) is inconsistent, rather than just inefficient, if heteroskedasticity is present, or if the errors are non-normal, in the present situation. Of course, inconsistency also arises,

with or without censoring or truncation, when various other assumptions are violated (such as those tested in (ii), (iii), (iv), (v), and (vii) of Example 3).

One of the attributes of random cell chi-square tests is their flexibility. This is illustrated here by the fact that each of the tests (i)-(ix) can be applied in the presence of censoring or truncation. All that is required is a straightforward adjustment of the cells to reflect the censoring or truncation. Figure 6 illustrates cell choices to test (iv) of Example 3 in the presence of censoring or truncation of Y_i at zero.[9] In both cases, the cells only partition the region where $Y_i \geq 0$. The normalized cell deviations can be used as in Example 3 to detect the direction of departure from the null hypothesis, if the test rejects.

5. Normal Linear Multivariate Regression, Seemingly Unrelated Regression and Simultaneous Equations Models--references: Malinvaud [29], Hausman [18]

For any of the three models considered here, one may be interested in any of the tests (i)-(ix) of Example 3 for some single equation in the multi-equation model. In the case of multivariate regression and seemingly unrelated regressions, exactly the same methods apply as in Example 3. In fact, the only difference between these cases and that of Example 3 is the method of parameter estimation.

For simultaneous equations, one can proceed exactly as in Example 3 to test (i)-(ix), except in those cases where a preliminary partition of $\mathcal{X}$ is made. In such cases, one has the option with a simultaneous equation of doing (1) a preliminary partition of $\mathcal{X}$ or (2) a preliminary partition of the space of covariates plus included endogenous variables. Either choice may be of interest when testing heterogeneity over the space of included

variables (ii), incorrect functional form (iii), or linearity of the effect of a single covariate (iv). Option (2) does not make sense in the case of a test of heteroskedasticity (vi). In carrying out a test of an omitted variable (v), the omitted variable could be either an exogenous or endogenous variable, without affecting the procedure outlined in Example 3. For testing exogeneity (vii), it only makes sense to consider exogeneity of one of the covariates in the equation. Finally, tests of normality (viii) and outlying errors (ix) can be treated identically in the simultaneous equations case, when one equation is of interest, as in Example 3.

Of special interest in multi-equations models is a test of the multivariate distribution of the errors. The quasi-maximum likelihood estimator (based on the assumption of multivariate normal errors) and numerous other common estimators are not asymptotically efficient in general, when the assumption of multivariate normal errors is violated. Further, the finite sample properties of most estimation procedures are much more fully understood with multivariate normal errors, than with other error distributions, due to the extensive array of exact results, see Phillips [43], and the wide variety of Monte Carlo results that utilize normal errors.[10] In consequence, a test of multivariate normality often is of interest.

One can test for multivariate normality by partitioning $\mathcal{Y} \times \mathcal{X}$ based on the values of residuals. Do not partition $\mathcal{X}$ first. It is natural to consider residual-based cells that are formed using ellipses. For testing multivariate normality of an observed vector (rather than unobserved errors), Moore and Stubblebine [36] consider cells whose boundaries are given by concentric ellipses centered at the sample mean, and with shapes determined by the sample covariance matrix. The same cell shapes can be used in the present context. These cells have power against peakedness, broad "shoulders," and

heavy tails. The normalized cell deviations can be used to determine which of these departures from normality is prevelant, if the test rejects. The cells used by Moore and Stubblebine do not have power against asymmetry. To enhance the test's power in this direction, one can partition the cells along the axes of the ellipse, see Figure 7. It may be desirable to choose the cells to be equiprobable or nearly so (see Andrews II, Section 5 for further comments on this issue), but perhaps with lower probability cells in the tails.

The above procedures can be extended straightforwardly when non-normal multivariate distributions are assumed for the errors. For example, Prucha and Kelijian [46] consider estimation of simultaneous equations with multivariate t-distributions for the errors.

To test for multivariate outliers in the errors, a similar procedure can be used as that above, but with the adjustments of (a) making the innermost cells larger and the cells in the tails more refined, and (b) estimating the parameters robustly.

6. Selection Models--reference: Maddala [28, Ch. 9]

The variety of different selection models is extremely wide. The simplest such model consists of a single linear regression equation and a single selection equation with binomial response variable. The latter equation determines when the dependent variable in the regression equation is censored. Various aspects of the selection equation can be tested in the manner indicated in Example 1 for categorical response variables. Alternatively, once the parameters of the regression equation have been estimated, any of the tests (i)-(ix) of Example 3 can be any carried out on the regression equation, in a manner analogous to that of Example 4 for censored regression.

In order to estimate selection models, one usually specifies a parametric

family for the joint distribution of the errors of the equations involved. If this parametric family does not include the true error distribution, then most estimation procedures are inconsistent. Hence, it is important to be able to test the distributional assumptions made on the errors.

We now suggest a method for constructing such a test, in the case of the simple selection model referred to above. Suppose the model is given by

$$Y_i = \begin{cases} X_{1i}'\beta + u_{1i} & \text{if} \quad I_i^* > 0 \\ 0 & \text{if} \quad I_i^* \leq 0 \end{cases},$$

$$I_i^* = X_{2i}'\alpha + u_{2i} ,$$

where I_i^* is an unobserved censoring variable, and (u_{1i}, u_{2i}) are errors that are assumed to have a bivariate normal distribution. Cells in $\mathcal{Y} \times \mathcal{X}$ are constructed based on the residuals $(\hat{u}_{1i}, \hat{u}_{2i}) \equiv (Y_i - X_{1i}'\hat{\beta}, I_i^* - X_{2i}'\hat{\alpha})$. The residual $\hat{u}_{2i}$ is not fully observed, however. It is only observed to be greater or less than $-X_{2i}'\hat{\alpha}$. Hence, the cells must be chosen so that only this information regarding u_{2i} is required. The cells of Figure 8 have this property. The heights of the horizontal edges of the cells are specified using the assumed (asymptotic) normal distribution of $\hat{u}_{1i}$ and its estimated mean and variance. The expected number of observations in each cell is given by the bivariate normal distribution using the estimated variance of u_{1i}, the estimated covariance of (u_{1i}, u_{2i}), and the normalized variance one of u_{2i}.

The choice of bivariate normal distribution in the above example is innocuous, a similar test with similar cell shapes can be constructed for any bivariate distribution.

7. Simultaneous Equations Models with Mixed Discrete, Continuous, Censored, and/or Truncated Endogenous Variables--reference: Maddala [28, Ch. 9]

The selection models of Example 6 are special cases of a class of simultaneous equations models with endogenous variables that may be discrete, continuous, censored, or truncated. Various aspects of these models can be tested using the approaches suggested in Example 5 for simultaneous equations models, coupled with the sort of adjustments for censoring and truncation that are illustrated in Example 4.

As in the case of selection models, the distributional assumptions placed on the errors are very important in these models, since their fulfillment is necessary for consistency for most estimators suggested in the literature. We illustrate how these assumptions can be tested in a simple two equation model with both endogenous variables censored. The model is

$$
\begin{aligned}
Y_{1i}^{*} &= \alpha_1 Y_{2i}^{*} + X_{1i}'\beta_1 + u_{1i} \\
Y_{2i}^{*} &= \alpha_2 Y_{1i}^{*} + X_{2i}'\beta_2 + u_{2i} ,
\end{aligned}
\tag{4.4}
$$

where (Y_{1i}^{*}, Y_{2i}^{*}) are latent endogenous variables and (u_{1i}, u_{2i}) are unobserved errors that are assumed to have a bivariate normal distribution. The observed variables are X_{1i}, X_{2i}, Y_{1i}, and Y_{2i}, where

$$
Y_{1i} = \begin{cases} Y_{1i}^{*} & \text{if } Y_{1i}^{*} \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad Y_{2i} = \begin{cases} Y_{2i}^{*} & \text{if } Y_{2i}^{*} \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{4.5}
$$

The parameters α_1, α_2, β_1, β_2, and the covariance matrix of (u_1, u_2) can be estimated by maximum likelihood, or by a two-stage procedure (see Maddala [28, Sec. 8.8]). Cells in $\mathcal{Y} \times \mathcal{X}$ are formed based on valued of the residuals $(\hat{u}_{1i}, \hat{u}_{2i}) \equiv (Y_{1i} - \hat{\alpha}_1 Y_{2i}^{*} - X_{1i}'\hat{\beta}_1, Y_{2i} - \hat{\alpha}_2 Y_{1i}^{*} - X_{2i}'\hat{\beta}_2)$. Both

residuals are observed only when Y_{1i}^* and Y_{2i}^* are positive. Hence, the cells are formed as those parts of concentric ellipses that lie in the region where the residuals are observed, see Figure 9. The expected cell frequencies are determined using the assumed bivariate normal distribution of the errors with estimated means and covariance matrix. Of course, different cell shapes in the upper right quadrant of Figure 9 are possible. Those chosen, however, are particularly convenient for detecting directions of departure from the null hypothesis (such as peakedness, broad shoulders, fat tails, or asymmetry), using the normalized cell deviations.

Note that the same method as above can be applied to any choice of bivariate distribution of the errors, not just bivariate normal errors.

8. Switching Regression Models--references: Maddala [28, Chs. 8, 9], Quandt and Ramsay [47], Gourieroux, Laffont, and Monfort [14, 15]

Here we consider switching regression models with or without endogenous switching. Most of the tests discussed in Example 3 for linear regression have analogues for switching models. As in Examples 6 and 7, the assumed error distributions must be correct for consistent estimation of the parameters. For example, consider the following endogenous switching model:

$$(4.6) \qquad Y_i = \begin{cases} X_{1i}'\beta_1 + u_{1i} & \text{if } I_i^* > 0 \\ X_{2i}'\beta_2 + u_{2i} & \text{if } I_i^* < 0, \end{cases}$$

$$I_i^* = X_{3i}'\alpha + u_{3i} ,$$

where I_i^* is a latent indicator variable, and the errors (u_{1i}, u_{3i}) and (u_{2i}, u_{3i}) are bivariate normal.

Two-stage estimation methods of this model (see [28]) start with estimation of α by maximum likelihood probit. The normality of u_{3i}, which justifies this technique, can be tested using the methods of Example 1. The second stage of such estimation methods uses the bivariate normality

of (u_{1i}, u_{3i}) and (u_{2i}, u_{3i}). These distributional assumptions can be jointly tested by partitioning $\mathcal{Y} \times \mathcal{X}$ based on the values of the residuals $\hat{u}_{1i}$ $(\equiv Y_i - X_{1i}'\hat{\beta}_1$ for $I_i^* > 0)$, $\hat{u}_{2i}$ $(\equiv Y_i - X_{2i}'\hat{\beta}_2$ for $I_i^* < 0)$, and $\hat{u}_{3i}$ $(\equiv I_i^* - X_{3i}'\hat{\alpha})$. $\hat{u}_{3i}$ is only observed to be greater than or less than $-X_{3i}'\hat{\alpha}$. $\hat{u}_{1i}$ and $\hat{u}_{2i}$ are only observed when $I_i^* > 0$ and $I_i^* < 0$, respectively. The cells illustrated in Figure 10 only require this information. Note that the cells to the right of the vertical line through $-X_{3i}'\hat{\alpha}$ correspond to values of $(\hat{u}_{1i}, \hat{u}_{3i})$, while the cells to the left of this line correspond to values of $(\hat{u}_{2i}, \hat{u}_{3i})$. To calculate the conditionally expected number of observations in each cell, one uses the bivariate distributions for (u_{1i}, u_{3i}) and (u_{2i}, u_{3i}) with estimated means, variances, and covariances.

9. Continuous Time Duration Models--references: Heckman and Singer [23, 24]

In duration models, the response variable Y_i is the length of time something lies in a given state. For example, it may be the length of time that a person is unemployed. For multi-state and multi-spell models, the response variable is vector-valued. In econometric applications, the distribution of Y_i is usually assumed to depend on a parametric linear combination of observed covariates X_i, and a vector of unobserved variables Z_i. That is, one specifies the parametric conditional distribution of Y_i given X_i and Z_i. In most cases, econometricians also specify a finite dimensional parametric family of marginal distributions for the unobservables, in order to facilitate estimation of the parameters of the conditional distributions of Y_i given X_i and Z_i. Heckman and Singer [23] argue that the choice of distributions for the unobservables is ad hoc. Further, they find that empirical results are often very sensitive to the choice of marginal distributions of the unobservables. In consequence, it would be useful to

be able to test the goodness-of-fit of the conditional distribution of Y_i given X_i, which is implied by the specified conditional distribution of Y_i given X_i and Z_i and the marginal distribution of Z_i.

Such a test can be constructed using random cell chi-square test statistics. Numerous methods of partitioning $\mathcal{Y} \times \mathcal{X}$ are possible. When interest is centered on the adequacy of the marginal distributions of Z_i, a natural procedure is to partition $\mathcal{X}$ based on values of $X_i'\hat{\beta}$ (where $X_i'\beta_0$ is the linear combination of covariates that affects the conditional distribution of Y_i given X_i and Z_i) and then to partition $\mathcal{Y}$, for each $\mathcal{X}$ cell, by one of the methods discussed in the examples above. For example, $\mathcal{Y}$ could be partitioned into equiprobable cells, perhaps with lower probability cells in the tail of the distribution.

If the specification of the conditional distribution of Y_i given X_i and Z_i also is suspect, then one may be interested in other tests, such as the analogues of tests of Examples 1 and 3 that are designed for general goodness-of-fit or for detecting heterogeneity over $\mathcal{X}$, omitted variables, or nonlinear functional form (of the affect of the covariates).

10. Panel Data Models--references: Chamberlain [7], Judge et al.[26]

The models considered here include most of the models discussed above, but the observations are assumed to be generated in a panel structure. Many of the same tests as above can be employed to investigate different aspects of the model in question. In addition, there are special aspects of panel data models that one may wish to test. For example, one may be interested in time homogeneity of the model. If a learning or start-up effect of a survey is of concern, one may want to test whether the model is the same for the first period, as for subsequent periods.

Consider a linear non-dynamic panel data model. Treat the observations

for a given individual over the entire time span as a single observation. A test of the above hypothesis can be constructed by estimating the parameters of the model using all observations but those of the first time period, and then partitioning $\mathcal{Y} \times \mathcal{X}$ based on variables corresponding to the first time period alone, using these parameter estimates. For example, one can partition $\mathcal{X}$ using only the covariates of the first time period, and then partition $\mathcal{Y}$, for each $\mathcal{X}$ cell, using the residuals from the first time period. A similar procedure can be used if there is censoring or if the response variable is categorical. Inspection of the normalized cell deviations provides information regarding the way in which the model differs between the first and subsequent periods, if the test rejects.

A second question of special interest in random effects panel data models is whether the random individual effects are independent of the covariates. To test this hypothesis in a normal linear random effects model, estimate the parameters using the within estimator, to ensure consistency of the estimator whether or not independence holds. Then, partition $\mathcal{X}$ based on covariates that are thought to be related to the random effect (perhaps, certain time invariant covariates), and partition $\mathcal{Y}$, for each $\mathcal{X}$ cell, based on the estimated random effects. The expected cell frequencies are calculated under the assumption of independence, using the normal distribution of the random effect. The normalized cell deviations give evidence of the nature of the association between the random effects and the covariates, in the case of rejection of independence.

11. Generalized Linear Models--references: Nelder and Wetterburn [37], McCullagh and Nelder [30]

In generalized linear models (GLMs), the response variable has an exponential family distribution (such as normal, Poisson, binomial, or gamma)

with parameters that are a known function of an unknown linear combination of covariates. The normal linear regression and the Poisson regression models are examples of GLMs.

A number of tests suggested in Example 3 for linear regression can be extended to GLMs. Test (i) for general goodness-of-fit can be extended, as can tests (ii)-(v), which investigate different aspects of the assumed linearity of the effect of the covariates on the response variable. To form cells for these tests when applied to GLMs, one can partition $\mathcal{X}$ by some method, as in Example 3, and then partition $\mathcal{Y}$, for each $\mathcal{X}$ cell, based on values of the response variable (rather than the residuals). For example, for each $\mathcal{X}$ cell, one might partition $\mathcal{Y}$ into equiprobable cells.

Alternatively, one might be interested in testing the adequacy of the assumed exponential family distribution, over the range of values that its parameter takes on (for different covariate values). This can be done by partitioning $\mathcal{X}$ according to values of $X_i'\hat{\beta}$, and then partitioning $\mathcal{Y}$ for each $\mathcal{X}$ cell. By partitioning $\mathcal{X}$ based on $X_i'\hat{\beta}$, one minimizes the within cell variation of the parameter values of the assumed exponential family distribution. By inspecting the normalized cell deviations, one can see the regions of $\mathcal{X}$ for which the exponential family distribution is adequate and those for which it is not adequate.

12. Nonlinear Regression and Simultaneous Equations Models With or Without Censoring and Truncation--reference: Amemiya [2]

The same tests as given above for linear regression and simultaneous equations models can be performed with nonlinear regression and simultaneous equations models (including implicit simultaneous equations models), except that cells cannot be based on residuals--at least without some assumptions on the form of nonlinearity.[11] This presents difficulties when trying to

construct tests for multivariate normality of the errors, but in other cases it can be circumvented. Instead of partitioning $\mathcal{Y}$, for a given $\mathcal{X}$ cell, based on residual values, one can partition $\mathcal{Y}$ using any method that gives straight edges to the cells. For example, one can linearize the curved regression surface (see Figure 11), or one can take edges perpendicular to the response variable Y_i (see Figure 12).

13. Frontier Production Function Models--references: Maddala [28, Ch. 6], Judge et al. [26, Ch. 20]

Frontier production function models often are used in measurement of technological inefficiency. These models take account of the fact that technological constraints impose an asymmetry on the distribution of errors in linear regression models for production functions. A number of alternative distributions have been suggested to replace the conventional normal error distribution. These alternatives include convolutions of normal distributions with half-normal distributions, and with exponential distributions.

The measure of inefficiency adopted for a given model depends on the assumed error distribution. Hence, it is important in these models to assess the adequacy of the assumed error distribution. This can be done, as in test (viii) of Example 3, by forming cells based on residual values. The shape of the cells can be designed specifically for the assumed error distribution, or the support of the distribution can be divided straightforwardly into equiprobable cells (perhaps with lower probability cells in the tails, see Andrews II, Section 5). The normalized cell deviations can be used to detect directions of departure from the null hypothesis, as discussed in Example 3.

5. CONCLUSION

This paper and Andrews II extend Pearson chi-square tests to models with covariates. The extension allows for data-dependent random cells, flexible cell shapes, and general methods of parameter estimation. These features yield great flexibility for the test. As illustrated by the examples above, one can choose random cells to test goodness-of-fit, or any of a wide variety of more specific aspects of a parametric model.

The introduction of random-cell chi-square tests is useful for three reasons. First, they provide tests in cases where no other tests are available. Second, since these tests are of a different nature than many existing tests, their power characteristics often are complementary to those of existing tests, in the sense of having maximum power against different alternatives than existing tests. This feature allows one to choose that test which is most powerful against alternatives of particular interest, in the common case where no uniformly most powerful test exists. Third, chi-square tests provide valuable direction of departure diagnostic statistics.

REFERENCES

[1] Amemiya, T.: "Qualitative Response Models: A Survey," Journal of Economic Literature, 19 (1981), 1483-1536.

[2] _____: "Nonlinear Regression Models, in Handbook of Econometrics, Volume 1, Chapter 6, ed. by Z. Griliches and M. D. Intrilligator. New York: North-Holland, 1983.

[3] _____: "Tobit Models: A Survey," Journal of Econometrics, 24 (1984), 3-61.

[4] Andrews, D. W. K.: "Random Cell Chi-Square Diagnostic Tests for Econometric Models: II. Theory," Cowles Foundation Discussion Paper No. 763, Yale University, New Haven, CT, 1985.

[5] Birch, M. W.: "A New Proof of the Fisher-Pearson Theorem," Annals of Mathematical Statistics, 35 (1964), 817-824.

[6] Cavanaugh, C. L.: "A Modified Multinomial Logit Model," manuscript, Department of Economics, Harvard University, 1985.

[7] Chamberlain, G.: "Panel Data," in Handbook of Econometrics, Volume 2, Chapter 22, ed. by Z. Griliches and M. D. Intrilligator. New York: North-Holland, 1984.

[8] Chernoff, H. and E. L. Lehmann: "The Use of Maximum Likelihood Estimates in χ^2 Tests for Goodness-of-Fit," Annals of Mathematical Statistics, 25 (1954), 579-586.

[9] Chibisov, D. M.: "Certain Chi-Square Type Tests for Continuous Distributions," Theory of Probability and Its Applications, 16 (1970), 1-22.

[10] Cramer, H.: Mathematical Methods of Statistics. Princeton: Princeton University Press, 1946.

[11] Dudley, R. M.: "Central Limit Theorems for Empirical Measures," Annals of Probability, 6 (1978), 899-929.

[12] Eisenhart, C.: "The Power Function of the χ^2 Test," Bulletin of the American Mathematical Society, 44 (1938), 32.

[13] Fisher, R. A.: "On the Interpretation of χ^2 from Contingency Tables and the Calculation of P," Journal of the Royal Statistical Society, 85 (1922), 87-94.

[14] Gourieroux, C., J. J. Laffont, and A. Monfort: "Coherency Conditions in Simultaneous Linear Equations Models with Endogenous Switching Regimes," Econometrica, 48 (1980), 675-695.

[15] _____: "Disequilibrium Econometrics in Simultaneous Equation Systems," Econometrica, 48 (1980), 75-96.

[16] Hammond, A.: "The Choosing of the 20," Science, Fifth Anniversary Issue, Vol. 5, No. 9 (1984), 9.

[17] Hartigan, J. A.: Clustering Algorithms. New York: Wiley, 1975.

[18] Hausman, J. A.: "Specification and Estimation of Simultaneous Equation Models," in Handbook of Econometrics, Volume I, Chapter 7, ed. by Z. Griliches and M. D. Intrilligator. New York: North-Holland, 1983.

[19] Hausman, J. A., B. H. Hall, and Z. Griliches: "Econometric Models for Count Data with an Application to the Patents-R&D Relationship," Econometrica, 52 (1984), 909-938.

[20] Hausman, J. A., B. D. Ostro, and D. A. Wise: "Air Pollution and Lost Work," manuscript, Department of Economics, M.I.T., Cambridge, MA, 1983.

[21] Hausman, J. A.,and D. A. Wise: "Stratification on Endogenous Variables and Estimation: The Gary Experiment," in Structural Analysis of Discrete Data: With Econometric Applications, ed. by C. Manski and D. McFadden. Cambridge, MA: M.I.T. Press, 1982.

[22] Heckman, J. J.: "The χ^2 Goodness of Fit Statistic for Models with Parameters Estimated from Microdata," Econometrica, 52 (1984), 1543-1547.

[23] Heckman, J. J.,and B. Singer: "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," Econometrica, 52 (1984), 271-320.

[24] _____: "Econometric Duration Analysis," Journal of Econometrics, 24 (1984), 63-132.

[25] Holland, P. W.: "Poisson Regression," manuscript, Educational Testing Service, Princeton, NJ, 1973.

[26] Judge, G. G., W. E. Griffiths, R. C. Hill, H. Lutkepohl, and T.-C. Lee: The Theory and Practice of Econometrics. Second Ed. New York: Wiley, 1984.

[27] Kendall, M. G. and A. Stuart: The Advanced Theory of Statistics, Volume II, Third Ed. New York: Hafner, 1973.

[28] Maddala, G. S.: Limited-Dependent and Qualitative Variables in Econometrics. Cambridge, Eng.: Cambridge University Press, 1983.

[29] Malinvaud, E.: Statistical Methods of Econometrics. Third Ed. New York: North-Holland, 1980.

[30] McCullagh, P. and J. A. Nelder: Generalized Linear Models. London: Chapman and Hall, 1983.

[31] McFadden, D.: "Conditional Logit Analysis of Qualitative Choice Behavior," in Frontiers of Econometrics, ed. by P. Zarembka. New York: Academic Press, 1974.

[32] _____: "Econometric Analysis of Qualitative Response Models," in Handbook of Econometrics, Volume II, Chapter 24, ed. by Z. Griliches and M. D. Intrilligator. New York: North-Holland, 1984.

[33] Moore, D. S.: "A Chi-square Statistic with Random Cell Boundaries," Annals of Mathematical Statistics, 42 (1971), 147-156.

[34] _____: "Generalized Inverses, Wald's Method, and the Construction of Chi-Squared Tests of Fit," Journal of the American Statistical Association, 72 (1977), 131-137.

[35] Moore, D.S. and M. C. Spruill: "Unified Large-sample Theory of General Chi-Squared Statistics for Tests of Fit," Annals of Statistics, 3 (1975), 599-616.

[36] Moore, D. S. and J. B. Stubblebine: "Chi-Squares Tests for Multivariate Normality with Application to Common Stock Prices," Communications in Statistics--Theory and Methods, A10(8) (1981), 713-738.

[37] Nelder, J. A. and R. W. N. Wedderburn: "Generalized Linear Models," Journal of the Royal Statistical Society, A135 (1972), 370-384.

[38] Nerlove, M. and S. J. Press: "Multivariate Log-Linear Probability Models in Econometrics," Discussion Paper No. 47, Center for Statistics and Probability, Northwestern University, Evanston, IL, 1980.

[39] Neyman, J.: "Contribution to the Theory of the χ^2 Test," Proceedings of the First Berkeley Symposium on Mathematical Statistics and Probability. Berkeley, CA: University of California Press, 1949, p. 239.

[40] Nikulin, M. S.: "Chi-Square Test for Continuous Distributions with Shift and Scale Parameters," Theory of Probability and Its Applications, 18 (1973), 559-568.

[41] Pagan, A. R.: "Model Evaluation by Variable Addition," in Econometrics and Quantitative Economics, ed. by D. F. Hendry and K. F. Wallis. Oxford: Basil Blackwell, 1984.

[42] Pearson, K.: "On the Criterion That a Given System of Deviations from the Probable in the Case of a Correlated System of Variables Is Such That It Can Be Reasonably Supposed to Have Arisen from Random Sampling," The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science, 50 (1900), 157-175.

[43] Phillips, P. C. B.: "Exact Small Sample Theory in the Simultaneous Equations Model," in Handbook of Econometrics, Volume 1, Chapter 8, ed. by Z. Griliches and M. D. Intrilligator. New York: North-Holland, 1983.

[44] Pollard, D.: "General Chi-Square Goodness-of-Fit Tests with Data-Dependent Cells," Z. Wahrscheinlichkeitstheorie verw. Gebiete, 50 (1979), 317-331.

[45] _____: "Strong Consistency of k-Means Clustering," Annals of Statistics, 9 (1981), 135-140.

[46] Prucha, I. R. and H. H. Kelejian: "The Structure of Simultaneous Equation Estimators: A Generalization Towards Nonnormal Disturbances," Econometrica, 52 (1984), 721-736.

[47] Quandt, R. E. and J. B. Ramsay: "Estimating Mixtures of Normal Distributions and Switching Regressions," Journal of the American Statistical Association, 73 (1978), 730-752.

[48] Rao, C. R.: Linear Statistical Inference and Its Applications. New York: Wiley, 1965.

[49] Rao, K. C. and D. S. Robson: "A Chi-Square Statistic for Goodness-of-Fit Tests Within the Exponential Family," Communications in Statistics, 3 (1974), 1139-1153.

[50] Romesburg, H. C.: Cluster Analysis for Researchers. New York: Van Nostrand Reinhold, 1984.

[51] Roy, A. R.: "On Chi-Squared Statistics with Variable Intervals," Technical Report, Department of Statistics, Stanford University, 1956.

[52] Späth, H.: Cluster Dissection and Analysis. New York: Wiley, 1985.

[53] Tauchen, G. E.: "Diagnostic Testing and Evaluation of Maximum Likelihood Models," Working Paper Number 84-01, Department of Economics, Duke University, 1984.

[54] Watson, G. S.: "The χ^2 Goodness-of-Fit Test for Normal Distributions," Biometrika, 44 (1957), 336-348.

[55] _____: "On Chi-Square Goodness-of-Fit Tests of Continuous Distributions," Journal of the Royal Statistical Society, Series B, 20 (1958), 44-61.

[56] _____: "Some Recent Results in Chi-Square Goodness-of-Fit Tests," Biometrics, 15 (1959), 440-468.

FOOTNOTES

[1]I would like to thank David Pollard and the participants of the econometrics workshops at Berkeley, Stanford, and Harvard/M.I.T. for their helpful comments and suggestions. The financial support of the National Science Foundation, through grant No. SES-8419789, is gratefully acknowledged. This paper was presented at the 1985 annual meetings of the Institute of Mathematical Statistics in Las Vegas.

[2]A difficulty arises with this approach if the integer values in the ratios defining the proportions in each strata are large relative to the sample size of interest. For example, if the observations are assumed to come from two strata in the ratio 67 to 33, and there are 200 underlying observations, then n is only equal to 2. Nevertheless, one would expect the asymptotics to "take hold" in this case for very small values of n.

[3]When no covariates are present, the conditional empirical process reduces to the standard empirical process, and so, tests based on either process are equivalent.

[4]By an asymptotically equivalent estimator, we mean an estimator that has the same linearized form (see equation (2.6) of Andrews II).

[5]Hammond [16], p. 9.

[6]The reduction of the degrees of freedom of the limiting chi-square distribution by one for each dimension of the parameter vector arises because the multinomial ML estimator is asymptotically equivalent to the estimator that minimizes $X_n^2(\theta)$ over $\theta \in \Theta$. The first order conditions for this minimization put L constraints on the partial derivatives of $X_n^2(\hat{\theta})$. Asymptotically, only the linearization of the L constraints are relevant. Hence, $X_n^2(\hat{\theta})$ behaves in large samples like a quadratic form in J independent normal variates conditional on the satisfaction of L linear constraints due to estimation, plus an additional constraint due to the fact that the variates must sum to zero. The quadratic form and linear constraints are such that the former has chi-square distribution with $J-1-L$ degrees of freedom. This heuristic explanation indicates the special character of the multinomial ML with regard to the effect of estimation on the limit distribution.

[7]Heckman assumes that the conditional distribution of the response variable given the covariates has a density. We presume this density is with respect to Lebesgue measure, otherwise his later statement regarding the degrees of freedom of the test statistic is not always correct.

[8]Heckman points out that the restriction to ML estimation can be relaxed by adjusting the test statistic appropriately. With some added complexity, due to additional terms in the weighting matrix, his asymptotic chi-square results still go through.

Also, note that the requirement that the cells be intervals is not very restrictive in Heckman's set-up, where the response variable is univariate and partitioning is based on that variable alone. If the response variable is allowed to be multivariate, or if partitioning can be based on both the response variables and the covariates, then the restriction to intervals or rectangles is restrictive.

[9]In the case of censoring, the cells that abut the X_i-axis have conditional probabilities that include the positive probabilities of outcomes occurring on the X_i-axis. In the case of truncation, the probability of an outcome on the X_i-axis is zero, and this is reflected in the cell probabilities. In consequence, if one desires roughly equiprobable cells, the pattern of cells has to be different in the censored and truncated cases.

[10]The same reasons can be used in Example 3 to motivate interest in the test (viii) of univariate normality of the errors.

[11]The reason is that with residual-based cells $\mathcal{C}$ is not necessarily a Vapnik-Červonenkis class, see Andrews II, Section 2.

[12]The cells in the 2-dimensional space of $(\hat{u}_{1i}, \hat{u}_{2i})$ depend on the subscript i through X_i, and hence, may appear to be different for different observations--a feature that is not allowed in the present framework. In the 4-dimensional space of $(Y_{1i}, Y_{2i}, X_{1i}, X_{2i})$, i.e., in $\mathcal{Y} \times \mathcal{X}$, however, the cells illustrated in Figure 8 correspond to cells that are independent of i, as required.

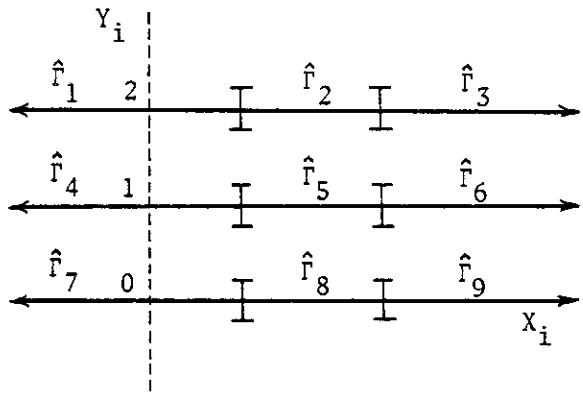

FIGURE 1--Categorical response model:
Cross-product cells

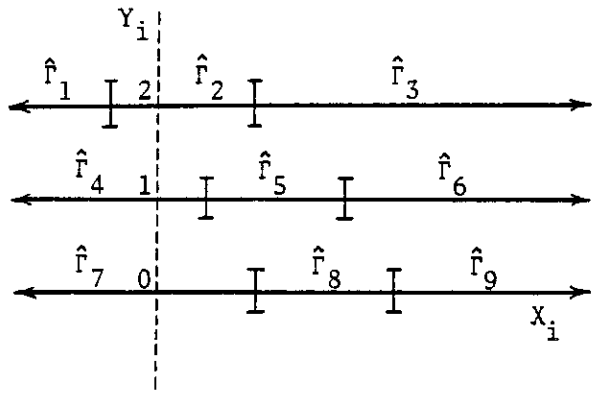

FIGURE 2--Categorical response model: $\mathcal{X}$ partitioned separately for each $\mathcal{X}$ cell

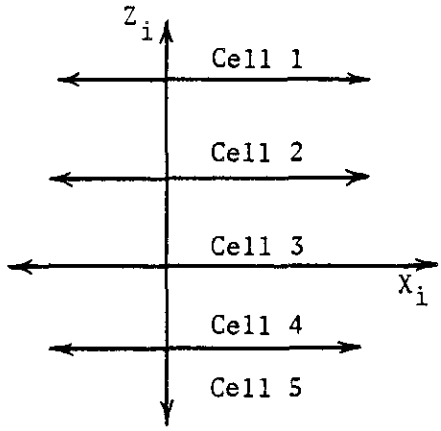

FIGURE 3--Partitioning of $\mathcal{X}$ based on omitted regressor variable Z

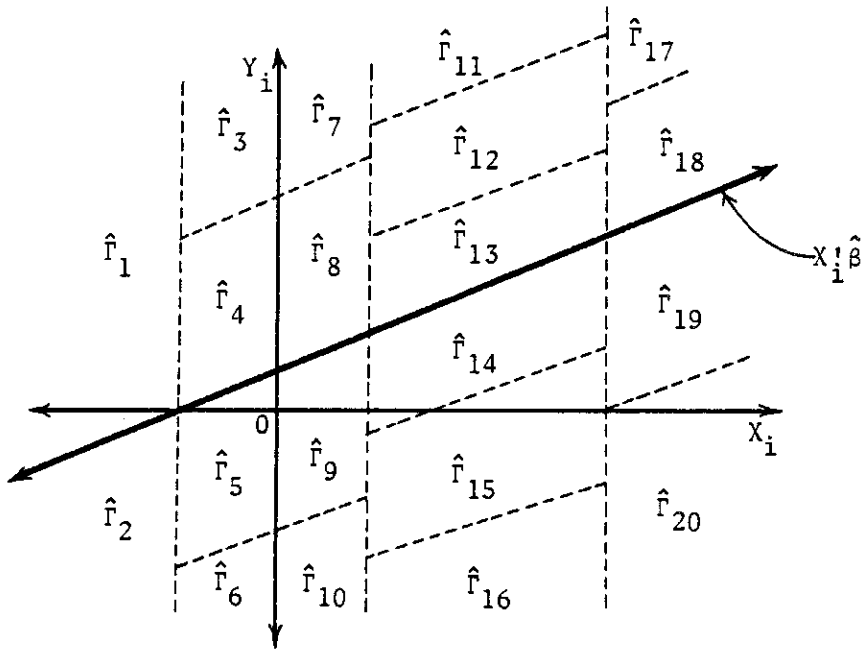

FIGURE 4--Linear regression:
Test of heterogeneity
over $\mathcal{X}$

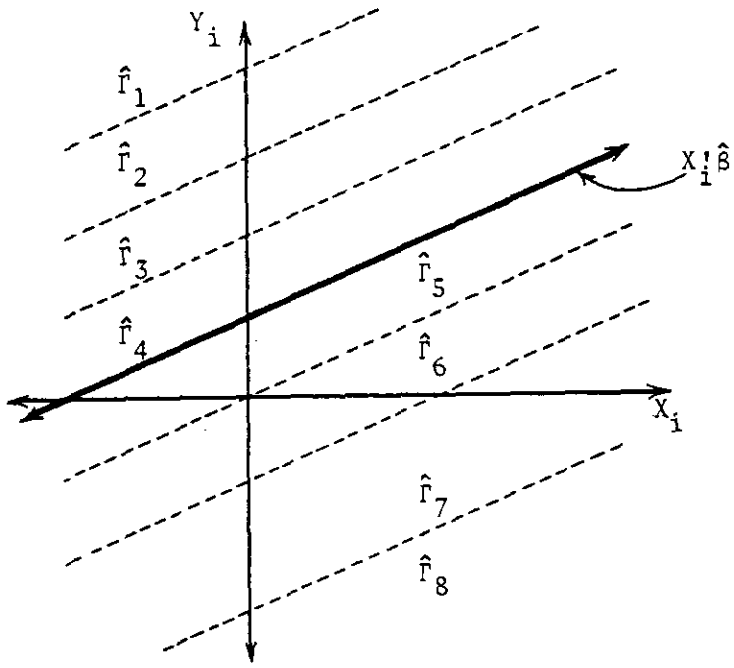

FIGURE 5--Linear regression:
Test of normality of
errors

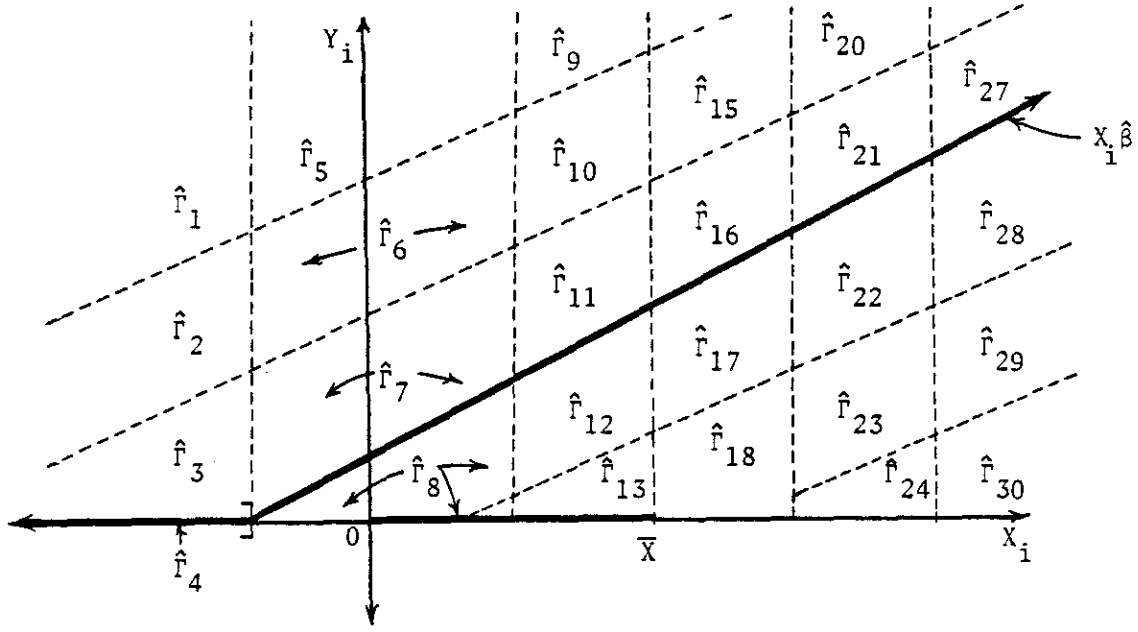

FIGURE 6--Censored or truncated linear regression: Test of linearity of the effect of a single covariate

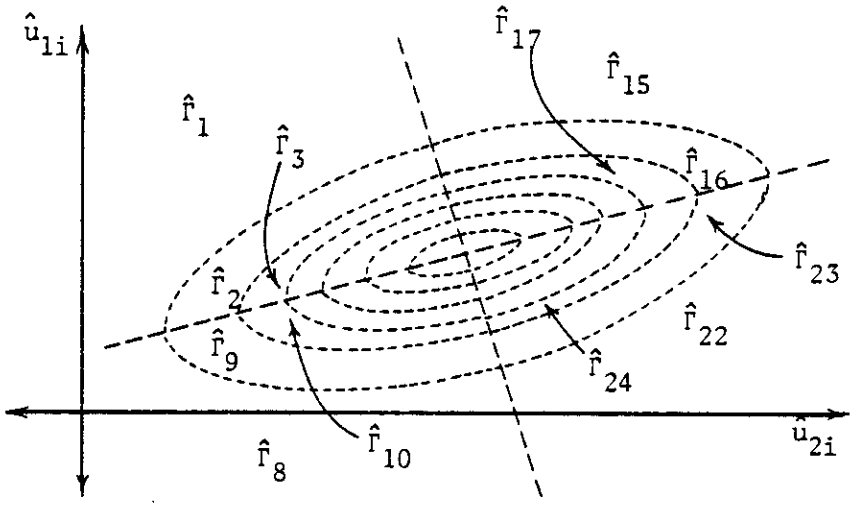

FIGURE 7--Test for multivariate normality based on residuals

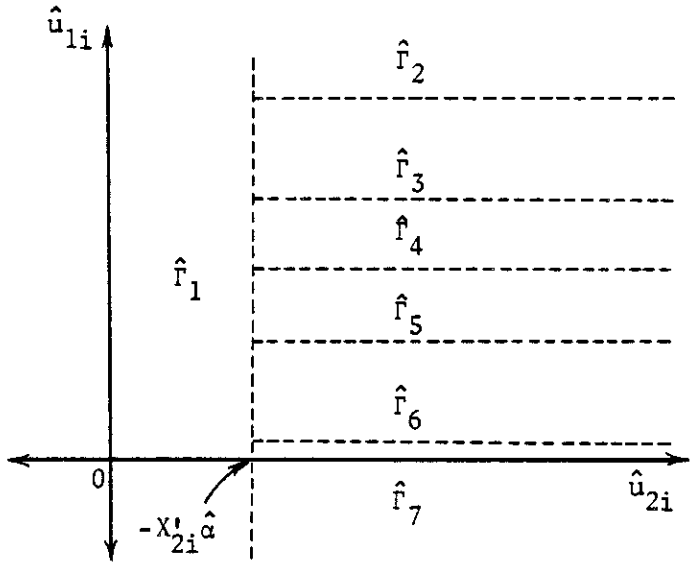

FIGURE 8--Selection model: Test for bivariate normality based on residuals[12]

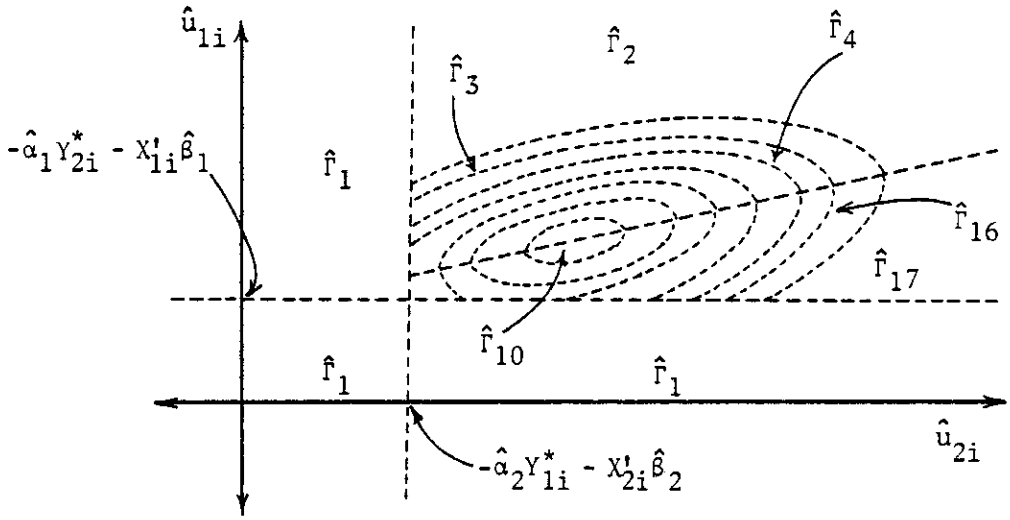

FIGURE 9--Censored simultaneous equations model: Test for bivariate normality based on residuals

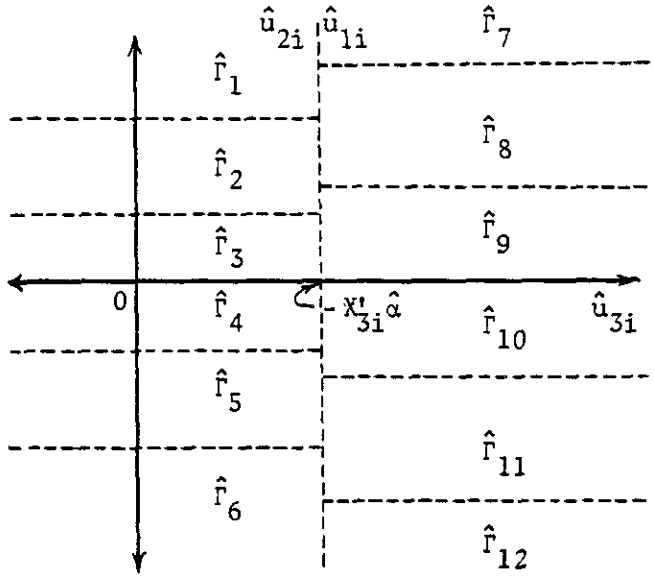

FIGURE 10--Switching regression model:
Test for normality errors

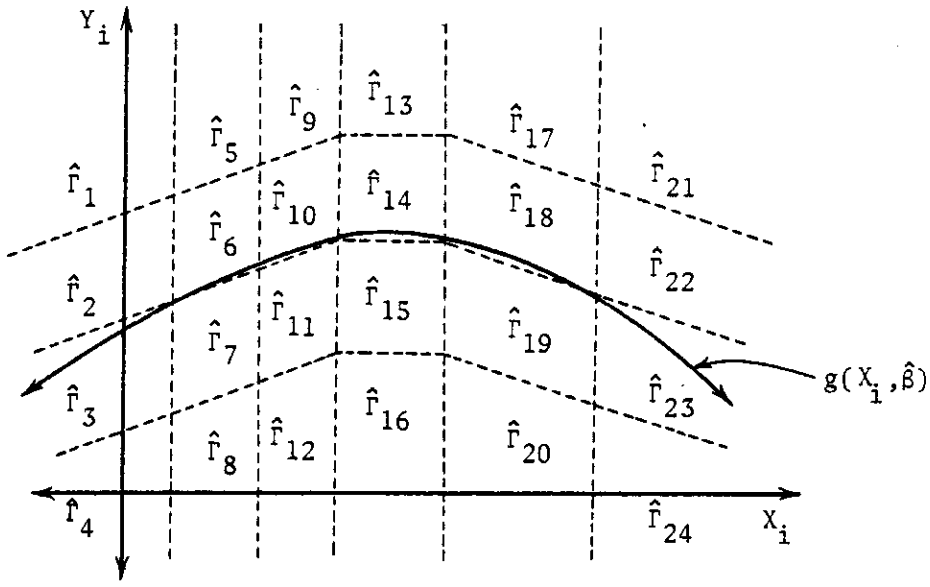

FIGURE 11--Nonlinear regression model:
Test for heterogeneity over $\mathcal{X}$

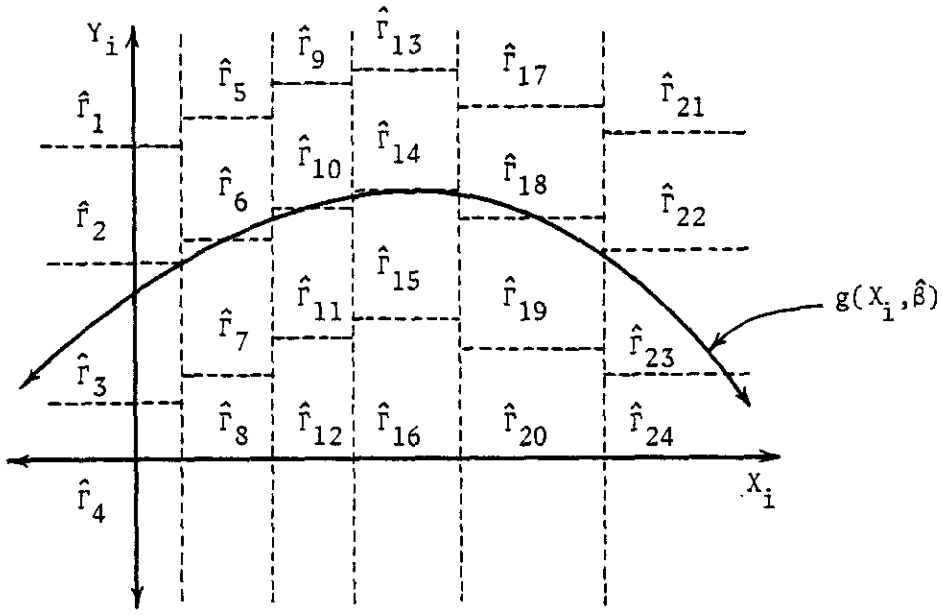

FIGURE 12--Nonlinear regression model:
Test for heterogeneity over $\mathcal{X}$